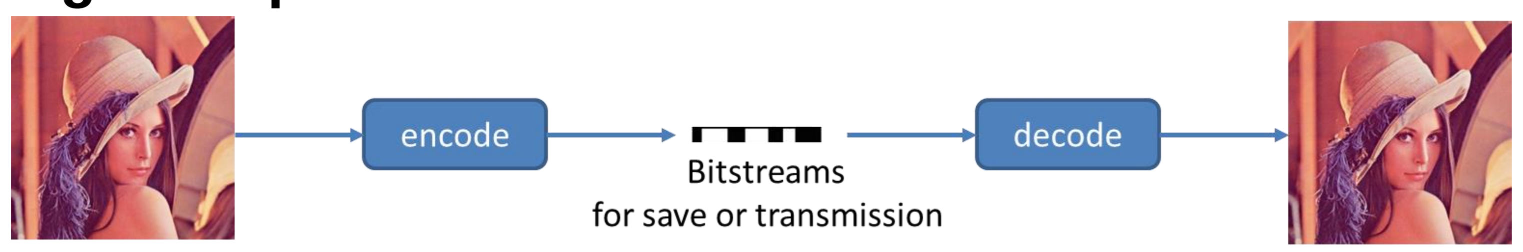
EVC: Towards Real-Time Neural Image Compression with Mask Decay

Guo-Hua Wang[1], Jiahao Li[2], Bin Li[2], Yan Lu[2]

[1]State Key Laboratory for Novel Software Technology, Nanjing University [2]Microsoft Research Asia

Introduction

Image compression:



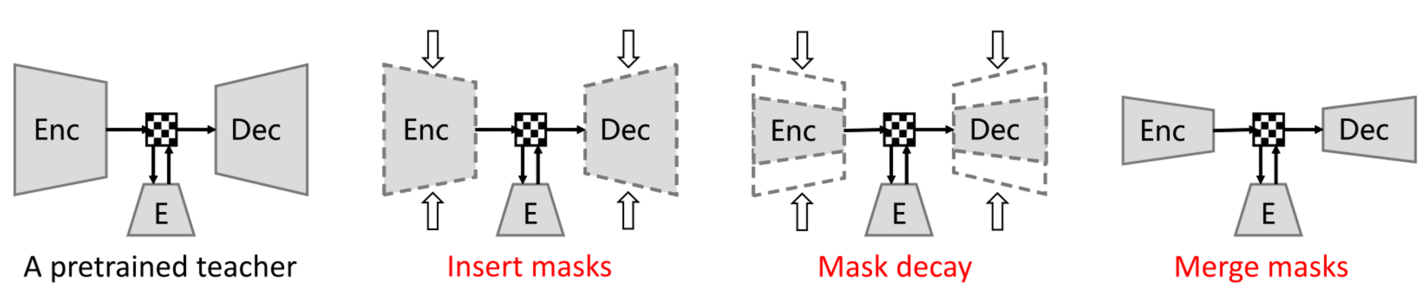
Main problems of neural image compression:

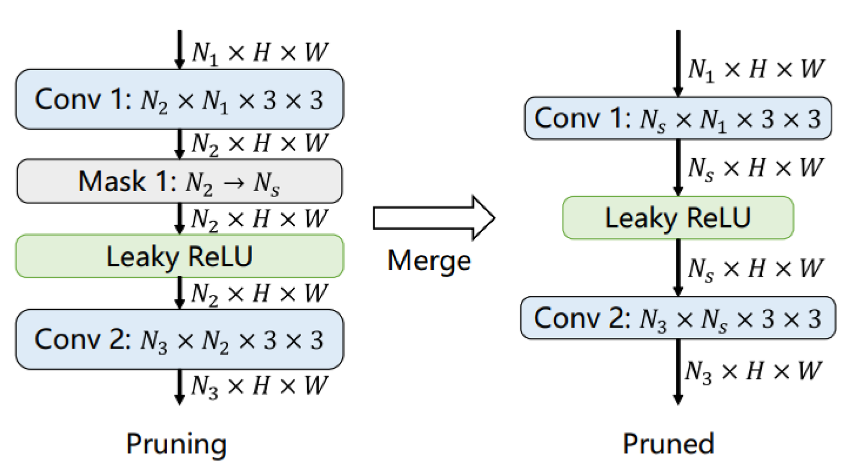
- The large complexity
- The inefficient rate-control

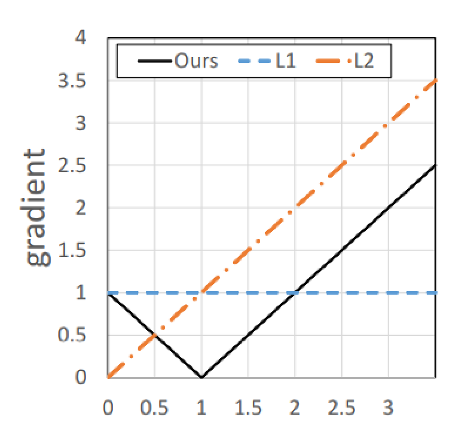
Our contributions:

- We propose an Efficient Variable-bit-rate Codec (EVC) for image compression. It enjoys only one model for different RD trade-offs. Our model is able to run at 30 FPS for the 768 × 512 inputs, while is on-par with other SOTA models for the RD performance. Our small model even achieves 30 FPS for the 1920 × 1080 inputs.
- We propose mask decay, an effective method to improve the student image compression model with the help of the teacher. A novel sparsity regularization loss is also introduced, which alleviates the shortcomings of L_p regularization. Thanks to mask decay, our medium and small models are significantly improved by 50% and 30%, respectively.
- We enable the encoding scalability for neural image compression. With residual representation learning and mask decay, our scalable encoder significantly narrows the performance gap from the teacher and achieves a superior RD performance than previous SlimCAE.

Mask Decay

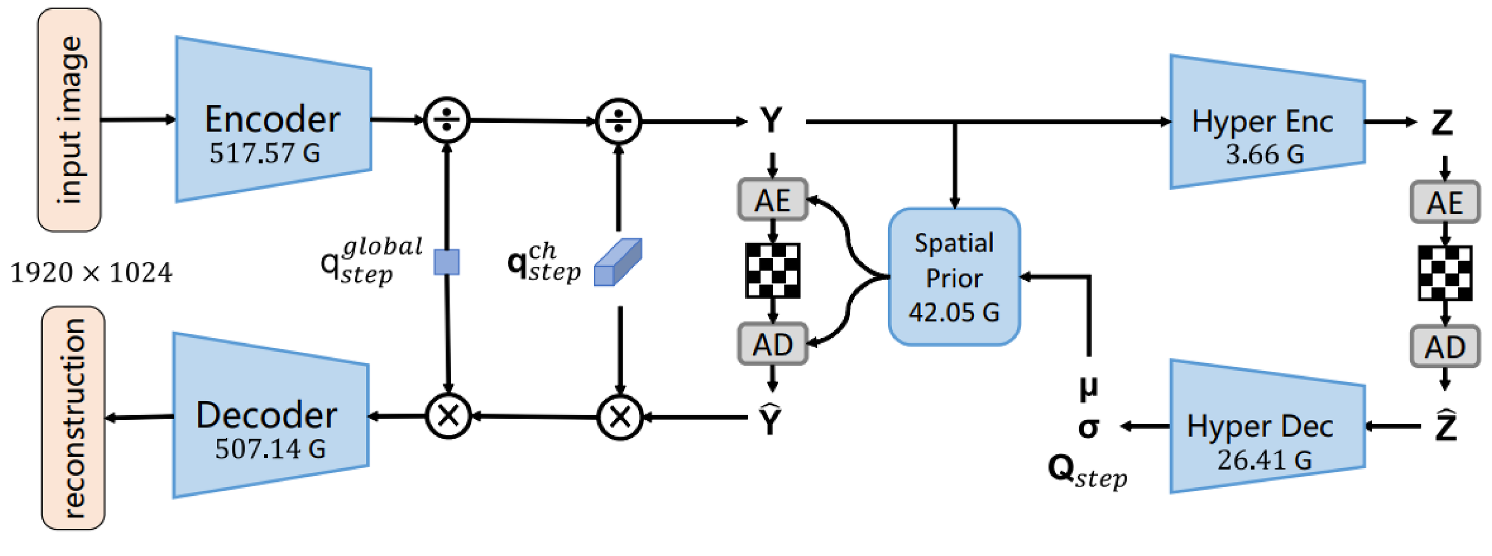


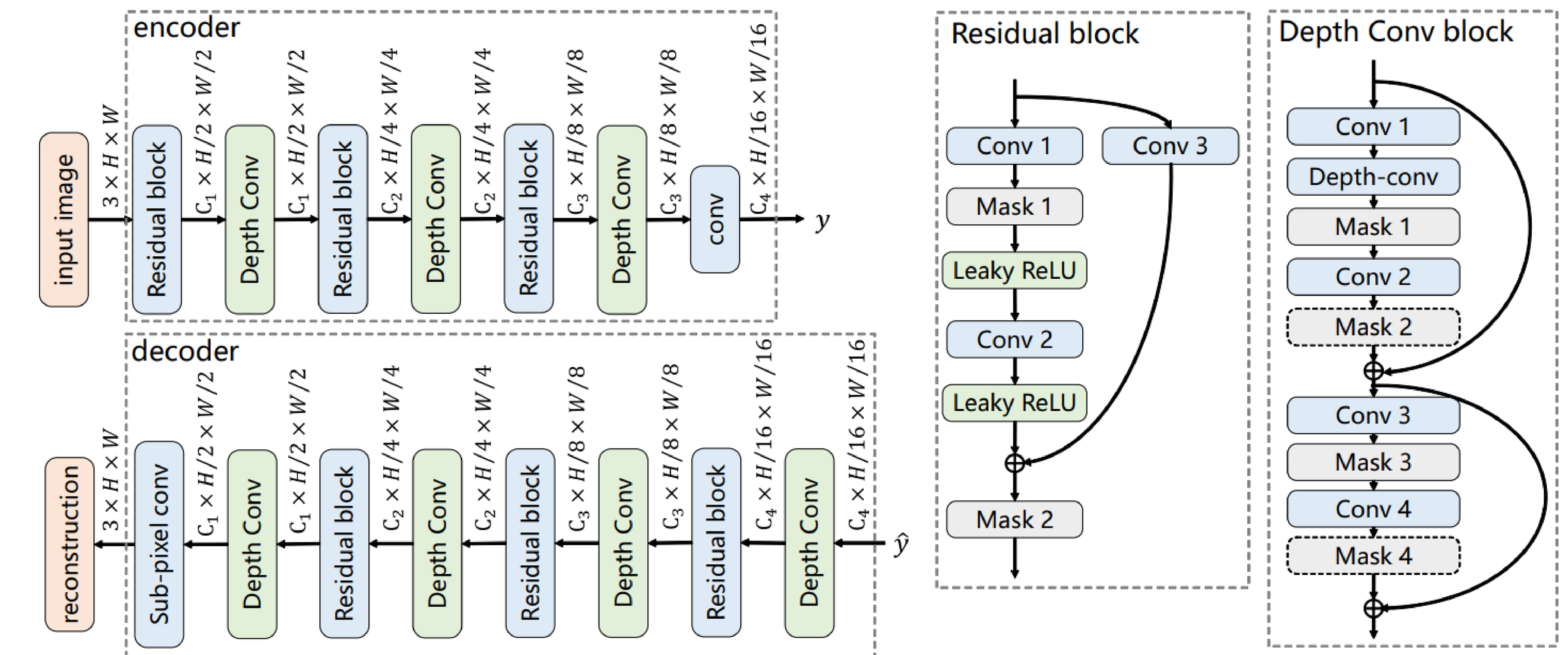




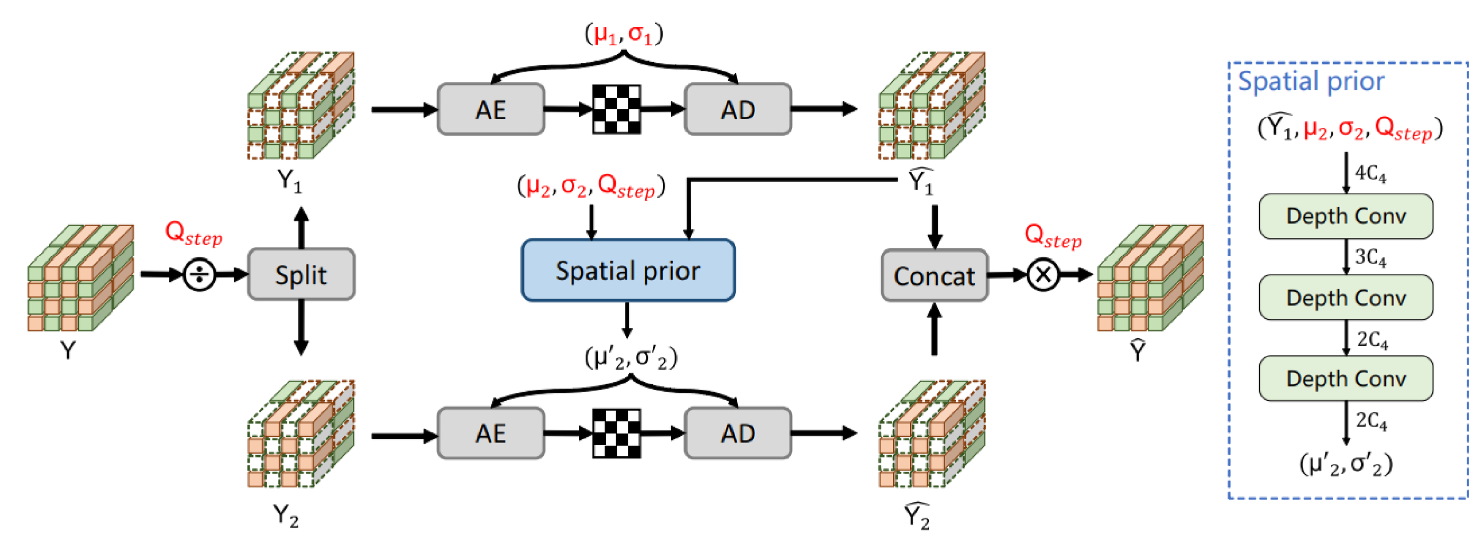
- The gradient of L2 vanishes when x approaches zero
- The gradient of L1 is a constant without considering its own magnitude
- Ours: $\frac{\partial \mathcal{L}_{sparse}(x)}{\partial x} = |x-1|, \quad \mathcal{L}_{sparse}(x) = \begin{cases} -\frac{1}{2}x^2 + x, & \text{if } 0 \le x \le 1, \\ \frac{1}{2}x^2 - x + 1, & \text{if } x > 1. \end{cases}$

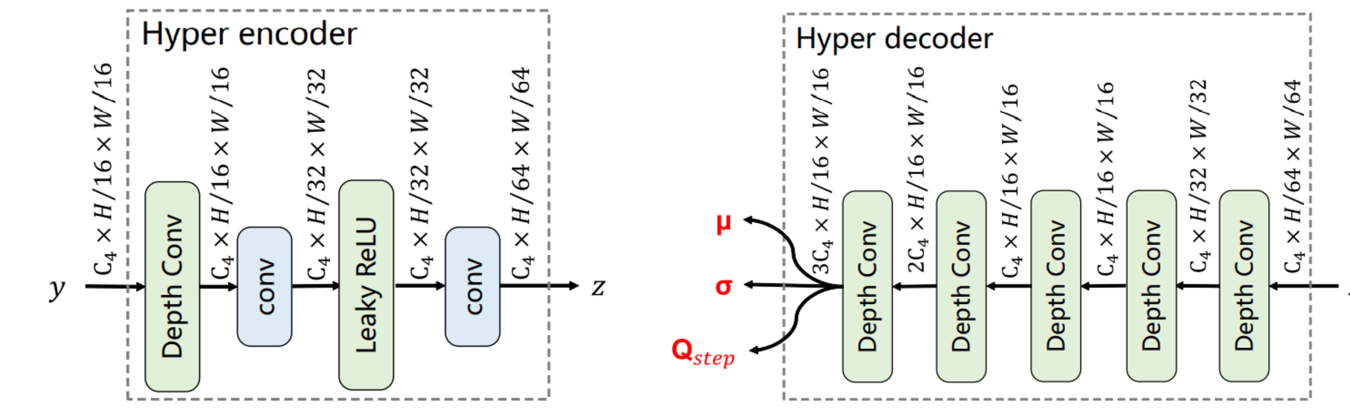
The EVC Framework



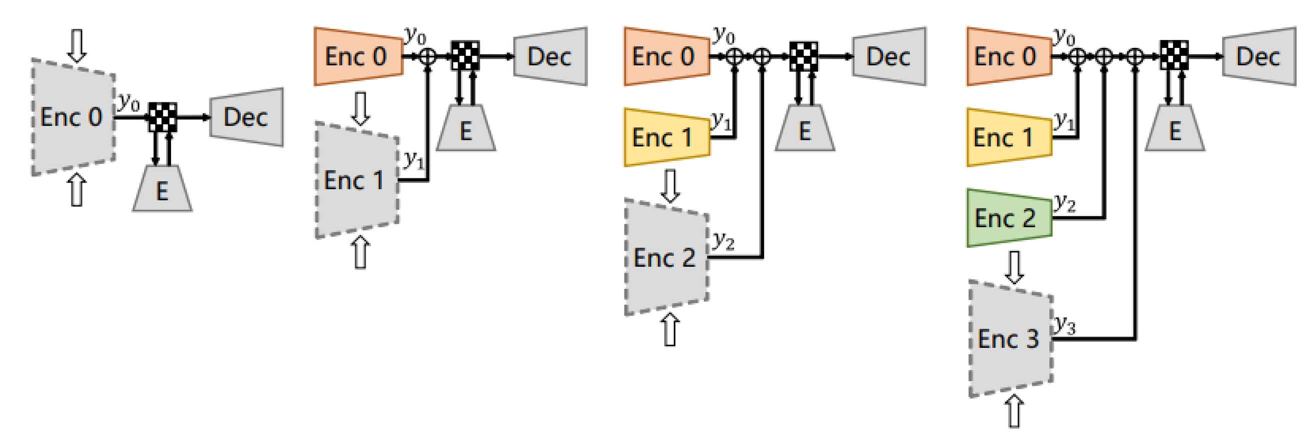


- adjustable quantization steps
- depth-wise convolution

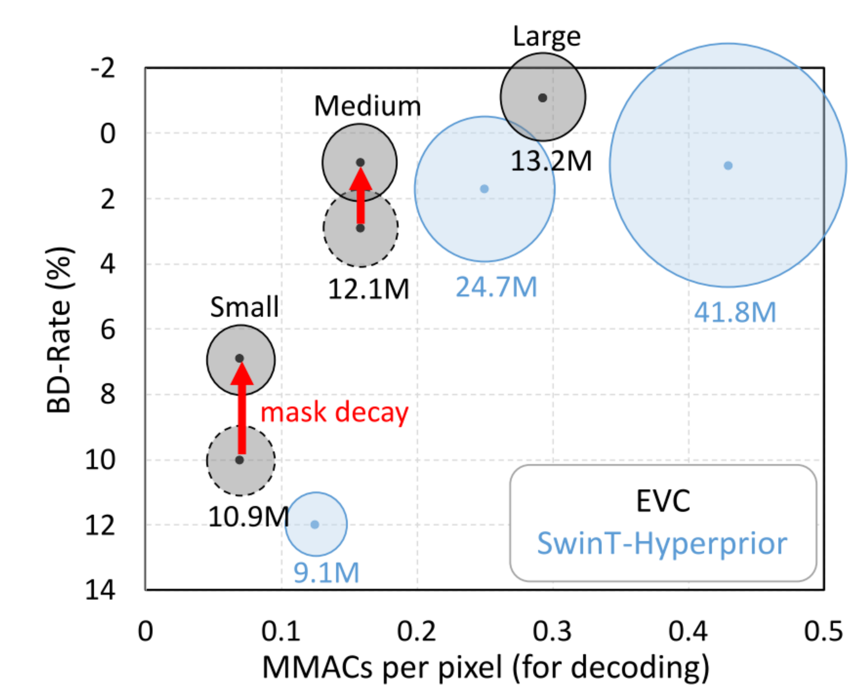


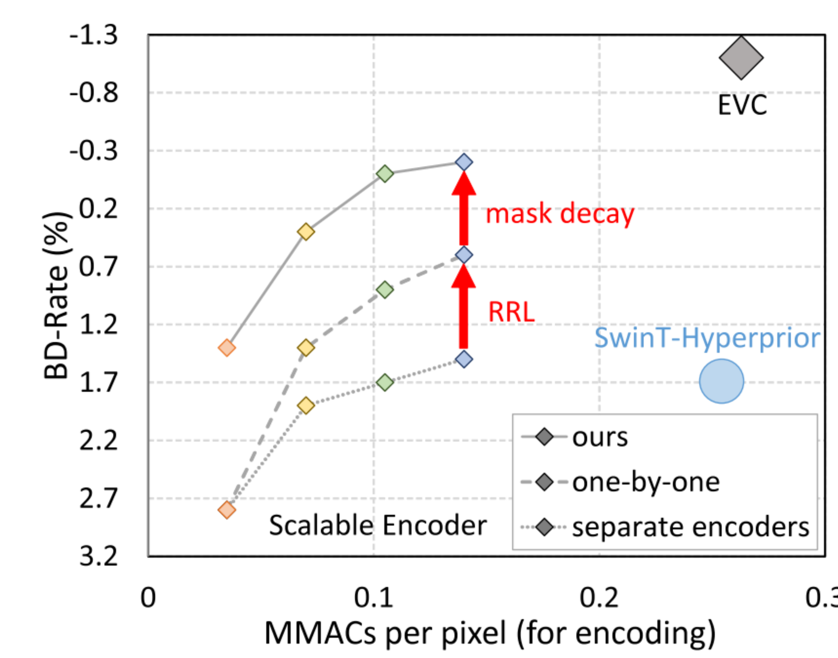


Residual Representation Learning (RRL)



Experimental Results





Latency

Resolution	GPU	Type	Entroformer	STF Transformer	STF CNN	EVC Large	EVC Medium	EVC Small
768 × 512	2080Ti	encoding	OM	176.3	158.5	63.0	44.7	**28.4**
		decoding	OM	202.3	210.2	41.1	**32.4**	**24.4**
	A100	encoding	816.8	115.9	96.4	**21.1**	**19.8**	**17.7**
		decoding	4361.9	143.2	118.0	**19.1**	**17.1**	**15.6**
1920 × 1080	2080Ti	encoding	OM	576.0	456.0	305.3	181.5	90.9
		decoding	OM	531.7	652.0	179.2	118.1	73.2
	A100	encoding	7757.4	355.6	278.1	84.2	56.3	**31.4**
		decoding	OM	354.8	281.7	60.2	46.5	**29.7**